Dan Stowell: Track record in research

Throughout my research so far (PhD and three years post-doctoral), my research vision has focused on innovative signal processing and machine learning within a context of applied work with sound. I have published on the core algorithmic engineering in the top journals and conferences, while also showing strength in cross-discplinary collaborations. I have also published highly-cited work on methods for evaluating machine listening systems within the context of use. I have recently been applying my research to natural sounds and birdsong, and based on this I have developed a strong vision for what can be achieved by developing my current strengths through this Early Career fellowship programme.

PhD

My aim for my PhD was to improve the integration of the human voice into digital music interactions. After exploring and evaluating existing machine-learning techniques, I introduced a new approach allowing automatic creation of audio "analogies", i.e. mappings from one sound domain to another (Stowell and Plumbley 2011). I also made general contributions to machine learning such as a novel efficient entropy estimation algorithm (Stowell and Plumbley 2009). Evaluating technologies which support creative expression is non-trivial: I introduced a novel methodology for rich evaluation of expressive music interfaces, subsequently adopted by others within the field (Stowell et al. 2009). Based on my PhD work I was selected as a finalist in the Guthman New Musical Instrument award 2010 (Georgia Tech University).

Throughout my PhD I made my software and data available under open licences. I became a lead developer of the "SuperCollider" open-source audio processing platform, building up its machine-listening capabilities and co-ordinating software releases via a global network of developers. I wrote a chapter in the official *SuperCollider Book* (MIT Press, 2012), and I also contracted a commercial software developer to adapt my research into an audio "app" for Android smartphones.

Post-doctoral

My first post-doctoral project was concerned with adapting music informatics technology for use in school music lessons. In order to ensure this research had real impact I designed and conducted an ethnographic study in two secondary schools (Stowell and Dixon 2011; Stowell and Dixon 2013). This study strongly informed my design of a new web service to find the chords played in YouTube videos <http://yanno.eecs.qmul.ac.uk/>. The service was extremely popular with the schools, achieving around 5000 video views per month soon after launch. I also integrated the service into the "Semantic Web" as Linked Open Data (RDF).

In 2012 I led a large team to host a complex multi-site conference and festival on the theme of music/audio computing ("The SuperCollider Symposium", budget £18,360, 116 participants). I led on grant applications, obtaining funding from the PRS Fund for New Music, Queen Mary University of London, and a private donor. I also led on a deliberately inclusive strategy, via low ticket prices, bursaries, a remix competition, and a free public exhibition in a park. We garnered a variety of media coverage (BBC, Daily Telegraph, and others), and our event was shortlisted for a prestigious Times Higher Award for Excellence and Innovation in the Arts.

In my most recent research I have been developing machine listening approaches for natural sounds including bird vocalisations. This began with a collaboration with biologists Briefer and McElligott within QMUL through which we produced a software tool to semi-automate the process of labelling birdsong audio. I then became interested in the many important open problems in computational bioacoustics for birds and other animals. I developed a novel graph-theoretic algorithm for tracking multiple birds in a sound scene (Stowell and Plumbley 2013), and collaborated with researchers at UPF (Barcelona) to combine this with high-resolution reassigned spectrograms for improved multiple bird tracking (Stowell et al 2013). As part of this theme I investigated a variety of probabilistic inference methods relating to the temporal evolution of sound, and also applied them to music and synthetic sounds (Stowell and Chew 2012, Stowell and Plumbley 2012).

I am co-organiser of an IEEE Challenge on Detection and Classification of Acoustic Scenes and Events, which aims to stimulate the research community to develop improved algorithms for general sound recognition in audio scenes. This successful initiative has attracted participation from 18 different research teams worldwide, and will have a presence at high-quality international conferences this year, in particular a special session at the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA).

In April this year I was invited to conduct a research visit to the group of Remi Gribonval at INRIA

(Rennes, France), experts on optimisation in signal processing. I collaborated with them to improve their *MPTK* sparse representation software, and to explore techniques such as chirplet ridge pursuit to achieve improved signal representations of birdsong.

In June this year I organised a one-day research workshop ("Listening in the Wild", budget £3,604) bringing together over 100 researchers in audio signal processing and in bird sound communication. This workshop was extremely successful, selling out quickly and with lively discussion at this cross-disciplinary interface. (Slides and videos: <http://c4dm.eecs.qmul.ac.uk/events/litw2013/>)

I have published a large amount of open code, open data and open-access research. This was recognised in June in the 2013 "Sound Software" Reproducible Research awards, in two prizes: I received an honourable mention for a conference paper with full code to reproduce the results; and I was co-author on a conference paper winning a prize for its associated open data and open-source code.

Leadership potential

Prior to my academic career, I led small teams of software developers and IT trainers to develop new tools and to ensure their successful take-up. Since then I have developed my leadership skills further through projects mentioned above: leading a large team of 9 organisers and 25 assistants to organise a complex multi-site conference and festival; conceiving and organising other research events; leading a project to adapt my PhD project into an Android app; supervising an MSc student as well as undergraduates performing data annotation; and taking the lead in development initiatives in various collaborative open-source software projects. My publication record also shows my leadership potential in research through a wide range of cross-disciplinary work shaped under my own initiative, drawing together different disciplines to realise new approaches to interaction with sound.

Public engagement

I have a solid track record of diverse public engagement activities, through appearances at science festivals, schools talks, and media appearances (BBC World Service, Guardian Science Podcast, Reuters TV News, BBC Technology News, BBC Focus Magazine, Wired, New Scientist, and more). In 2009 I was accepted onto the EPSRC "NOISEmakers" scheme and worked with the EPSRC on a public engagement activities over a number of years, interacting with hundreds of children and adults about science and engineering.

Collaborators

David Clayton is an international research leader in songbird ethology (behaviour) and genetics. Clayton's work led to the discovery of genes that play an active role in vocal communication. In 2003, he organized a broad set of international collaborations which led to the complete sequencing of the songbird (zebra finch) genome.

Richard Turner is a Lecturer in the Computational & Biological Learning Lab (University of Cambridge). His research lies at the interface between computer perception, neuroscience and machine-learning.

Marc Naguib leads the Behavioural Ecology Group at Wageningen University (Wageningen, Netherlands). He has published extensively on bird communication in noise for over 2 decades.

Thierry Aubin leads the Bioacoustics Team at Université Paris Sud (Orsay, France). His research prioritises a multidisciplinary approach combining tools from signal processing, ethology, neurobiology and ecology, with the aim to understand the acoustic communication systems of animals.

Structured machine listening for soundscapes with multiple birds

Case for support

In this Early Career fellowship I will establish a world-leading capability in automatic inference about songbird communications via novel *machine listening* methods, working collaboratively with experts in machine listening and with experts in bird behaviour and communication. Automatic analysis has already shown benefit to researchers in efficiently characterising recorded bird sounds, but there are still many limitations in applicability. The techniques developed will specifically be designed to handle noisy multi-source audio recordings, and to infer not just the presence of birds but the structure of the signals and the interactions between them. Such methods will be a leap beyond the current state of the art in bioacoustics, allowing researchers to study not just sounds recorded in the lab under controlled conditions, but also field recordings and archive recordings found in public audio archives.

Importantly, not only will I apply modern signal processing and machine learning techniques, but I will also develop new techniques inspired by this application area. This fellowship is not about contributing from one field to another, but about building up UK research strength in this cross-discplinary research topic. In order to make the most of this possibility, I will host research workshops and an open data contest to serve as focal points for research attention, and I will also conduct a public engagement initiative to engage the widest possible enthusiasm for this exciting field of possibility.

My aim will be achieved through a series of objectives to be realised in collaboration with researchers in signal-processing, machine learning, bioacoustics and ethology. I elaborate on the objectives in *Programme and Methodology* below; but first, some context.

Background

The subject matter of this research is the vocal sound made by *songbirds*. True songbirds (members of the suborder *Passeri*) have an important distinction from the majority of the other birds that we hear around us: juvenile songbirds go through stages of *vocal learning* in which they learn the detailed structure of song from their father and from others singing nearby (Marler and Slabbekoorn 2004). This phenomenon gives songbirds high scientific importance: songbird vocal learning is one of the most direct evolutionary and developmental parallels with human vocal learning, meaning that songbirds are model organisms for many current lines of research in neurology, ethology, genetics and linguistics (Marler and Slabbekoorn 2004, Chapter 8; Clayton 2013; Abe and Watanabe 2011). A recent landmark was the first full sequencing of the genome of a songbird, the zebra finch (Warren et al. 2010), which paves the way for understanding of the neurology and the rich evolved variation in songbird communication patterns.

Note that in formal studies there is a distinction between two types of bird vocalisation: *song* is often relatively long and complex, used as display for mating and territorial defence, while *calls* are short sounds used for purposes such as warning of predators or keeping members of a flock in contact (Marler and Slabbekoorn 2004, esp. chapter 5). This proposal concerns both song and calls, since both contain implicit information about bird identities and social structures that are amenable to analysis.

The study of sound in animal communication (*bioacoustics*) has traditionally involved a large degree of manual sound analysis such as inspection of spectrograms, but since around the turn of the century there has been increasing interest in automated procedures, which bring benefits such as repeatability and scalability. A notable example is Tchernichovski's work on measuring song similarity (Tchernichovski et al. 2000), which has been used in various studies of song development and cultural transmission (Lipkind and Tchernichovski 2011). Such methods are applied to individual birds recorded in low-noise laboratory conditions, and the techniques do not easily apply to field recordings or less-controlled archive recordings, or even to laboratory recordings of multiple birds interacting together.

There is a body of work which studies animal communication in less controlled conditions, though it commonly relies on more manual analysis techniques. Topics of study include songbirds deliberately overlapping each other to signal aggressive rivalry (Naguib 2010), or avoiding overlap when singing as part of a dawn chorus (Malavasi 2013); mechanisms used to distinguish conspecific song from other sounds (Aubin and Bremond 1983); and the connections between the social structure of a bird colony and the dynamics of vocal exchanges (Elie et al 2011).

Behavioural studies aside, bird sounds are also important for monitoring purposes. Birds are often detectable by sound at least as much as by appearance, especially in woodland, and so bioacoustics is of growing importance in monitoring bird population distributions, and how they vary over time and with the seasons (Laiolo 2010). Bird population monitoring is important not just for its own sake, but also as an indicator of environmental change (Morecroft 2013). Autonomous acoustic monitoring is increasingly

recognised as an important tool for these purposes, with current concern over whether automatic systems can match the sensitivity and robustness achievable with human observers (Laiolo 2010; Digby 2013).

Various groups have studied automatic detection of birds in sound; a notable example is that of Briggs et al. (2012) who go beyond the common single-item-classification paradigm, describing a system that can detect multiple bird species in a scene. A useful snapshot of the state of the art is provided by the online challenge run by the French "SABIOD" project (led by Hervé Glotin) as part of the International Conference on Machine Learning 2013 <http://sabiod.univ-tln.fr/icml2013/>. This event challenged researchers to develop algorithms to determine which bird species were present in a recording. Results were encouraging, with detection rates approaching 70% (by the "area under the curve" measure) across 35 candidate species. There is much scope for improvement, but further: the leading submissions in that competition largely used relatively standard machine-learning with no temporal sequence modelling. These methods thus do not offer any "parsing" of the audio scene that might be used for animal communications studies.

There is therefore still a significant gap between the characterisation of individual sounds, as performed manually or automatically in controlled conditions for animal communications studies, and the type of algorithmic analysis that can be applied to field recordings, archive recordings, or even lab recordings of multiple birds. In my recent work I have made some contributions toward bridging this gap, most notably a technique for automatically tracking multiple birds of the same species through an audio scene (Stowell and Plumbley 2013; Stowell et al. 2013). This and related methods (Barker et al. 2005; Mahler 2007) use probabilistic models of a multi-source scene to make it possible to provide structured inference about the scene, with background noise and multiple sources explicitly accounted for. This avenue is thus a promising route for taking the automated analysis of bird communication to the point where it can make inferences about communication networks in field recordings.

A significant gap in the state of the art in bird sound analysis is that the signal representations most used are standard representations such as short-time Fourier transform magnitudes ("spectrograms"), mel-frequency cepstral coefficients (MFCCs), or linear prediction coefficients (LPCs). These techniques are widely used and understood, but unfortunately since they assume signals are *pseudo-stationary* they can omit or obscure information that is highly relevant in bird vocalisations. In particular, details of the fine temporal structure such as rapid frequency modulation (FM) are extremely important in songbirds: not only are songbird vocal mechanisms specifically evolved to produce rapid FM (Goller 2012), but songbirds can perceive fine detail of FM and it influences their behavioural responses (Lohr 2006; de Kort 2009). Rates of FM up to 100 kHz/s are common. However, such FM information is poorly captured in pseudo-stationary representations, providing an impoverished signal for downstream analysis. This motivates developing and evaluating methods based on nonstationary signal processing. I have demonstrated in recent work that chirplet analysis of birdsong can lead to improved recognition (Stowell and Plumbley 2012b) and improved tracking (Stowell et al. 2013), and I have further results in preparation.

There are various signal-processing techniques which merit further study in regard to temporal fine structure and FM. A general area of current interest is that of *sparse representations*, in which we seek a representation which is sparse in some domain, which can yield powerful inference (Fevotte et al 2008; Plumbley et al 2010; Xu et al. 2013). Another approach which offers the prospect of useful structured information is model-based probabilistic inference such as that of Turner and Sahani (2012), which starts from a perceptually-inspired model to infer FM detail and other modulation features in a sound signal, using the raw audio signal as input rather than preprocessed features. Their method treats a sound as a single entity rather than decomposing it into presumed component sources, but it is a useful illustration of the possibility that model-based probabilistic inference could be used to unify what are commonly treated as two separate steps: signal representation and inference about the scene.

Host institution

The Centre for Digital Music (C4DM), at Queen Mary, University of London (QMUL) is a world-leading multidisciplinary research group in the field of Digital Music & Audio Technology. In the past year I have spent time at various institutions through collaborations with researchers from UPF Barcelona, Oxford, UCL, INRIA Rennes, and I was even offered a job by one. However, I consider the C4DM the ideal host for the project I propose. The key is the C4DM's depth of research strength in both intelligent audio processing and machine learning, combined, which will provide a richer substrate for the various aspects of the project than would a more traditional computer science or bioacoustics group. The C4DM has over 60 full-time members, and research funding since 2007 totals over £17m, from EPSRC, EU, Royal Society, Leverhulme Trust, TSB, JISC, Mellon Foundation and industry sources. The group has developed several

robust technologies for music and audio research, including Sonic Visualiser, an open-source cross-platform framework for analysis of music and audio, downloaded over 200,000 times since 2007.

In addition to this, an opportunity has arisen in that David Clayton, an international research leader in songbird ethology (behaviour) and genetics (see *Track Record*), has recently moved his lab from the USA to QMUL. In Work Package 2 in this fellowship I will collaborate with the Clayton group, making use of their zebra finch facilities, and Clayton will provide mentoring to ensure I develop appropriate connections through my research with ethology and the biological sciences.

Technical expertise in the C4DM includes topics which directly support the themes in this fellowship, as evidenced by current funded projects. For example, EPSRC project "Semantic Media" (Sandler and Kudumakis, £572,750) brings many researchers together to address the challenge of the navigation of time-based media collections. EPSRC leadership fellowship "Machine Listening using Sparse Representations" (Plumbley, £1,236,776) extends the state of the art in sparse representations for signal processing, and uses it to address machine listening tasks.

The C4DM uses the IT infrastructure of the School of Electronic Engineering and Computer Science (EECS), and additionally owns dedicated state-of-the-art high-performance computer clusters with ample data storage capacities available to all researchers. Through the EPSRC grant for Software Sustainability the Centre employs two full-time software engineers to maintain existing research program code, and to support the software implementation of novel methods. The C4DM also has a suite of professional sound-proof rooms for listening experiments and recordings. It has hosted many international conferences well-known in the field (DAFx, ISMIR, ICAD, ICA, HCI, AES, MPEG). The School also hosts the Computer Science for Fun (*cs4fn*) project, with whom I will work on public engagement (see *Pathways to Impact*).

Programme and methodology

The overall methodology for this fellowship is to develop signal processing and machine learning (SP/ML) techniques applied to recorded sounds containing bird vocalisations. I have found in various recent evaluations (Stowell and Plumbley 2012b, 2013; Stowell et al 2013) that existing techniques in SP/ML can fall short of being suited to the task, due to assumptions such as pseudo-stationarity or single-source (see *Background* above). Hence novel variants of processing chains or novel probabilistic models are often needed in order to fit the task properly, and these will be my focus. Specific innovations that will be needed include the generalisation of existing inference models to handle multi-species scenarios or to infer extra parameters such as bird personality traits, connected to aspects of zoological interest. I will also develop my ideas around integrating sparse signal representations with larger-scale probabilistic inference, in liaison with sparse representations experts in my host research group, and also specifically explore connections with Turner and Sahani's work on inferring soundscape parameters and frequency modulations (2011/2012), through a research visit with Turner.

In order to tailor my work so that the outputs have strong potential impact in the biological sciences, I will collaborate with biological experts throughout to develop techniques relevant to ideas and hypotheses in the field. In particular, in collaboration with the Clayton group I will record zebra finches in small social "colonies" in the lab. This is designed to be a staging post between the more common single-individual recording and the complexities of natural field recordings. I will also work with wider datasets already collected by others, including public sound archives (Cornell, Xeno-Canto, Berlin) as well as those of the British Library Sound Archive, and those recorded by collaborators such as the Naguib and Aubin groups.

Looking more widely, my ambition is not just to conduct the work but also to foster the growth of the research community in this cross-disciplinary topic. I will do this in particular by hosting research workshops, as well as through collaborations and research visits.

Project scoping: The focus of this fellowship is on songbirds, because of their zoological and ecological importance. I will maintain awareness of research in other creatures (bats, insects, mammals) and follow-on research may lead to collaborations in those areas. Songbirds are an important object of study, and the zebra finch is a model organism in that field, hence my specific focus on that in the work with the Clayton group. However, I will not restrict attention purely on zebra finches, in order to enable wide applicability and to allow alternative lines of enquiry.

Also, note that some research projects make use of spatial information when tracking birds, for example with multiple microphones spread through a forest. However, spatial information is not highlighted in the work plan, for the important reason that the vast majority of recordings of interest are mono or uncalibrated stereo: this is not just the case for legacy audio in archives, but also for modern recordings made by amateurs, and also by professionals with portable recording rigs.

Work Package 1 (WP1): Modelling techniques (months 1-24)

The focus for the first WP will be to explore the nonstationary signal processing and machine learning techniques needed to characterise birds singing and calling in a multi-source environment.

I have recently demonstrated the value of nonstationary SP methods (see *Background* above) for birdsong. Therefore I will start by evaluating a range of nonstationary SP methods for their suitability for use in later WPs. Specific methods include: varieties of chirplet analysis; sparse representations with dictionary learning; and sparse representations with parametric dictionaries. I will collaborate with colleagues in my host research group who have diverse experience in this field.

Machine learning starting points will include the novel inference method I recently introduced, *multiple Markov renewal process* (MMRP) inference. I will study extensions to this such as the incorporation of interactions between calling birds. Other starting points of particular pertinence are multi-source tracking models such as the *probability hypothesis density filter (PHD filter)* (Stowell and Plumbley 2012a), and the work of Richard Turner on probabilistic scene analysis and amplitude and frequency demodulation (Turner and Sahani 2012). Regarding this last point I have arranged to conduct a research visit to Cambridge Machine Learning Group (University of Cambridge, UK), to collaborate with Turner on development of these probabilistic inferences about audio scenes. I will also develop some ideas in combining the different stages of processing into a unified inference directly from audio, as is done in Turner's approach but for single sounds only.

Outcomes: Published evaluations of signal processing and machine learning techniques for bird vocalisations; novel methods or novel variants of appropriate methods, including unified inference.

WP2: Zebra finch case studies (months 3-18; collaboration with the Clayton lab in QMUL)

WP2 will be conducted in collaboration with Clayton's research group based in the School of Biological and Chemical Sciences at my host institution (QMUL). The Clayton group conducts behavioural, neuroscientific and genetic studies using the zebra finch as a model, and they will provide access to zebra finches reared as part of their research stock. Our first activity, after obtaining ethical approval for studies with live birds, will be to create a small recorded "pilot" dataset by acoustic and video monitoring of a small colony of 20 zebra finches, in a laboratory setting. This is designed to create a dataset of intermediate difficulty, in that we will set up a social scenario with multiple interacting birds, but without some of the difficulties of outdoor field recordings such as noise, distance and interference. This data will be augmented by annotations of the activity of the zebra finches by paid annotators.

Then, using the work from WP1 we will analyse the pilot data, adapting techniques as necessary. This will involve further study of zebra finch sounds in particular, adapting the signal analysis to ensure that it is well-suited to the vocal ranges and variations of the species.

The zebra finch data will also feed into a public data challenge later in the fellowship (see WP5). The data will be published online as open data for others to use. A portion may be held back as private data for evaluating algorithms submitted to the challenge, as is common in such challenges.

Following this first round of data collection and analysis, we will have an informed perspective from which to conduct further studies. The study design will be developed in light of the pilot: it will involve a further round of zebra finch recordings, potentially in different social configurations, and will be designed to evaluate experimentally the ability of algorithms to capture behaviourally relevant issues such as bird personality differences and interactions between and within sexes.

Outcomes: Public dataset of zebra finches recorded in multi-bird social interactions; improvements to techniques of WP1; publications arising from collaborative study with Clayton lab.

WP3: Further modelling, communications networks (months 25-39)

After having conducted the relatively controlled work with zebra finches, in WP3 I will broaden the perspective, extending the multi-source tracking of WP1 to analyse larger multi-species communications networks, and to incorporate richer social interactions such as group responses to predators. In order to develop this I have arranged two research visits to centres of expertise in bird communication:

* Marc Naguib and his group at the Behavioural Ecology Group at Wageningen University (Wageningen, Netherlands). We will study data collected by the group on interactions among great tits as well as other species, in order to generalise my audio analysis methods to various bird communications networks.

* Thierry Aubin and his Bioacoustics Team at Université Paris Sud (Orsay, France). We will study the connections between audio modelling and communication structures via data about birds in different types of habitat (e.g. wrens in forests vs. skylarks in open environments).

Outcomes: Published collaborations with bioacoustics research groups; improved/generalised methods; studies of their application in songbirds more generally than zebra finches.

WP4: Applications in audio archives (months 40-48)
One large and untapped resource for the study of bird communication is generic audio archives, not just those specifically designed to collect birdsong. There are important challenges in making this possible: some challenges are the same as those addressed in evaluations throughout the fellowship (noise robustness, scalability) but also further challenges including the lack of calibration and the interference from strong foreground sounds such as voice. The British Library Sound Archive (based in London) is an archive of world importance, and holds many collections which can be used for this topic. I have a relationship with the BLSA from previous work with its head Richard Ranft, and will make use of their archive in particular. I will also conduct studies with free public audio databases such as FreeSound.

Outcomes: Published research on discovering bird communications in audio archives; improved scalability and robustness of algorithms.

WP5: Building the research community (months 13-60)
Part of the aim of this fellowship is to build up a strong research community in this area. As a focus of this community-building, I will host a biannual research workshop, which will be modelled on the successful workshop I hosted this year ("Listening in the Wild"). This will not be the only venue in which machine listening researchers and bioacousticians may encounter one another (see for example the bioacoustics workshop at the 2013 International Conference on Machine Learning), but it will help to build up a research network in the UK and Europe, and will have a specific focus on themes such as multi-source sound analysis and connections with animal perception.

In the run-up to Listening in the Wild 2017, I will host a public "data challenge" based on data recorded in WP1. This will challenge researchers to design algorithms capable of answering questions pertinent to ethologists, such as "which zebra finch is speaking?" or "is this a shy or a dominant bird?" The approach for this challenge will build on my recent experience co-organising an IEEE data challenge for sound recognition. I will also organise a related special session at an international conference.

Outcomes: Research workshops in 2015 and 2017; public data challenge.

WP6: Team development and skills (months 1-60)
I have experience in team leadership and in training others (see *Track Record*), and will use this Early Career fellowship to develop my abilities and position myself for leading a research group in future. I will supervise Masters student projects (QMUL MSc course *Digital Signal Processing*) each year. I will attend further training for PhD supervision, then I will recruit and supervise a PhD student over the final years of the project on the topic of *"Probabilistic models of inter-species bird acoustic communication"*, which will connect with the work of WP3 in particular. Also, to fill gaps in my skills in bird recognition by sight and sound, I will attend a short residential course provided by the British Trust for Ornithology (BTO).

Outcomes: Improved skills; successfully completed student projects.

WP7: Public engagement (months 13-60)
In Year 2, timed to come after the initial development work and the first research workshop, I will work with the host institution's *cs4fn* team to produce an issue of the QMUL "Audio!" schools magazine on the topic of bird communication and machine listening. "Audio!" is a mature route to spread ideas of audio engineering and machine listening into thousands of schools. Connected with this, I will also carry out talks in schools.

In Year 4 I will develop a larger project: an interactive exhibition piece which explores ideas about the structure of bird song and calls and the use of computational techniques to model sonic interactions, in an entertaining and accessible format. For this I will build on my extensive experience in artistic public engagement (see Track Record). This will be a portable exhibit that can be toured around the UK and exhibited both in art galleries and municipal venues (e.g. public libraries or parks), in order to promote public engagement more broadly than the main metropolitan centres. Opportunities will be sought to present the work both on its own and as part of wider exhibitions, such as events curated by the Sonic Arts Network and the Royal Society.

Outcomes: Edition of schools magazine; talks in schools; exhibit shown around the UK; media appearances.

WP8: Extension research (months 49-60)
In the final WP I will integrate the latest research results into my own work, conduct small pilot studies to

explore potential next directions, and develop follow-on project proposals and collaborations.

Outcomes: Follow-on project proposals and new collaborations.

Risks and mitigation

Risk	Probability	Severity	Mitigation
Problems in algorithm performance	Med	Med	I have already introduced and evaluated various methods, so there are many fallback routes to follow if the more ambitious problem formulations present insurmountable difficulties.
Project partners unavailable	Low	Med	Other partners would be contacted.
Zf colony unavailable	Low	Med	Existing data would be used, and I would seek other collaborators who could help provide access to zebra finches.
Zf studies unproductive	Low	Med	I will not focus solely on zebra finches but also other species (e.g. tits in Naguib's lab; wrens, skylarks in Aubin's lab).
Techniques inapplicable to problems in biology	Low	Med	I will have an internationally leading biologist as a mentor/advisor (Clayton), and will work with his lab and with other biologists throughout the fellowship to ensure relevance in biological disciplines.
BL data unavailable	Med	Low	There are public-domain datasets available with fewer copyright restrictions (e.g. FreeSound, Xeno-Canto) that would provide similar subject matter albeit with lesser UK importance.
Public engagement exhibit problems	Low	Low	I have strong track record in public engagement, and in running public exhibited works, and I will also work with colleagues in the host institution who have this experience (e.g. *cs4fn*).

Importance

National importance: This proposal works across disciplines in which the UK has acknowledged strengths, but relatively little cross-disciplinary work of the type I propose. Yet the UK has a tradition of birdwatching and caring for birds, as evidenced by the millions of amateur members of the RSPB and BTO, and the UK also has international strength in machine learning and machine listening for other types of sound: for example, my host institution's recognised leading position in music audio analysis. Hence there is fertile ground for the UK to build an international leadership in machine listening for birds.

Societal challenges: Fluctuations in bird population/migration are indicators of environmental change (Morecroft 2013), and autonomous acoustic monitoring is increasingly recognised as an important tool for monitoring these (Laiolo 2010, Digby 2013). In the UK, woodland birds have declined to about 80% of their level in the early 1970s, and farmland birds to around 50% of their level; 52 birds in the UK are currently Red Listed as Birds of Conservation Concern (RSPB and others 2013). Acoustic monitoring is especially appropriate for birds since in many cases they are more often heard than seen. This fellowship thus is nationally important in developing our research base in this area. It will also facilitate large-scale analysis of bird sounds in developmental studies, in which birdsong is an important model with parallels to human language (Marler and Slabbekoorn 2004, chapter 8). A further societal challenge is that of unlocking value from large archives such as the British Library Sound Archive, a collection of international significance.

EPSRC priorities: The EPSRC's overriding priority for ICT fellowships is *Working Together*, to which this fellowship aligns strongly. The core aim is to develop probabilistic machine listening within the engineering contexts of signal processing and machine learning, while directly working with biologists to ensure that the techniques developed will benefit data-driven research in bioacoustics and beyond. Further, the topic lies within research areas that EPSRC prioritises: *Digital signal processing* and *Statistics and applied probability* are current targets for growth, while *Music and acoustic technology* and *Artificial intelligence technology* are targets for maintenance at current levels.

Academic importance: please see the *Academic Beneficiaries* section in the form.

UK economy and industry: this fellowship will build a thriving research focus in the UK that enables

industrial applications such as: autonomous monitoring systems; computer systems with context-sensitive audio interfaces (e.g. in mobile phones); and source separation for hearing aids and cochlear implants.

References cited

K. Abe and D. Watanabe. Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience*, 14:1067–1074, 2011.

T. Aubin and J. C. Bremond. The process of species-specific song recognition in the skylark alauda arvensis. an experimental study by means of synthesis. *Zeitschrift fur Tierpsychologie*, 61(2):141–152, 1983.

J. P. Barker, M. P. Cooke, and D. P. W. Ellis. Decoding speech in the presence of other sources. *Speech Communication*, 45(1):5–25, 2005.

F. Briggs, B. Lakshminarayanan, L. Neal, X. Z. Fern, R. Raich, S. J. K. Hadley, A. S. Hadley, and M. G. Betts. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *J Acoustical Society of America*, 131:4640–4650, 2012.

D. F. Clayton. Genomics of memory and learning in songbirds. *Annual Review of Genomics and Human Genetics*, 14(1), 2013.

D. F. Clayton, C. N. Balakrishnan, and S. E. London. Integrating genomes, brain and behavior in the study of songbirds. *Current Biology*, 19(18):R865– R873, 2009.

S. R. de Kort, E. R. B. Eldermire, S. Valderrama, C. A. Botero, and S. L. Vehrencamp. Trill consistency is an age-related assessment signal in banded wrens. *Proc. Royal Society B: Biological Sciences*, 276(1665): 2315–2321, 2009.

A. Digby, M. Towsey, B. D. Bell, and P. D. Teal. A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods in Ecology and Evolution*, 4(7):675–683, 2013 .

J. E. Elie, H. A. Soula, N. Mathevon, and C. Vignal. Dynamics of communal vocalizations in a social songbird, the zebra finch (Taeniopygia guttata). *Journal of the Acoustical Society of America*, 129:4037, 2011.

C. Fevotte, B. Torresani, L. Daudet, and S.J. Godsill. Sparse linear regression with structured priors and application to denoising of musical audio. *IEEE Trans. Audio, Speech, and Language Processing*, 16(1):174–185, 2008.

F. Goller and T. Riede. Integrative physiology of fundamental frequency control in birds. *Journal of Physiology-Paris*, 2012.

P. Laiolo. The emerging significance of bioacoustics in animal species conservation. *Biological Conservation*, 143(7):1635–1645, 2010.

D. Lipkind and O. Tchernichovski. Quantification of developmental birdsong learning from the subsyllabic scale to cultural evolution. *Proceedings of the National Academy of Sciences*, 2011.

B. Lohr, R. J. Dooling, S. Bartone, et al. The discrimination of temporal fine structure in call-like harmonic sounds by birds. *J Comparative Psychology*, 120(3):239–251, 2006.

R. P. S. Mahler. *Statistical Multisource-Multitarget Information Fusion*. Artech House, Boston/London, 2007.

R. Malavasi and A. Farina. Neighbours' talk: interspecific choruses among songbirds. *Bioacoustics*, 22:33–48, 2013.

P. R. Marler and H. Slabbekoorn. *Nature's music: the science of birdsong*. Academic Press, Massach., USA, 2004.

M. Morecroft and L. Speakman. *Terrestrial biodiversity climate change impacts summary report*. Technical report, Living With Environmental Change Partnership.

M. Naguib and D. J. Mennill. The signal value of birdsong: empirical evidence suggests song overlapping is a signal. *Animal Behaviour*, 80(3):e11, 2010.

M. D. Plumbley, T. Blumensath, L. Daudet, R. Gribonval, and M. E. Davies. Sparse representations in audio and music: From coding to source separation. *Proceedings of the IEEE*, 98(6):995–1005, 2010.

RSPB and 24 other UK organisations. The State of Nature 2013. http://rspb.org.uk/ourwork/science/stateofnature/

D. Stowell and M. D. Plumbley. Multi-target pitch tracking of vibrato sources in noise using the GM-PHD filter. In *Proc 5th Intl Workshop on Machine Learning and Music (MML12)*, 2012a.

D. Stowell and M. D. Plumbley. Framewise heterodyne chirp analysis of birdsong. In *Proceedings of EUSPICO*, 2012b.

D. Stowell and M. D. Plumbley. Segregating event streams and noise with a Markov renewal process model. *Journal of Machine Learning Research*, 14, 1891—1916, 2013.

D. Stowell, S. Musevic, J. Bonada, and M. D. Plumbley. Improved multiple birdsong tracking with distribution derivative method and Markov renewal process clustering. In *Proc Int Conf Audio and Acoustic Signal Processing (ICASSP)*, 2013.

O. Tchernichovski, F. Nottebohm, and others. A procedure for an automated measurement of song similarity. *Animal Behaviour*, 59(6):1167–1176, 2000.

R. E. Turner and M. Sahani. Decomposing signals into a sum of amplitude and frequency modulated sinusoids using probabilistic inference. In *Proc Int.Conf Acoustics, Speech and Signal Processing (ICASSP)*, pages 2173–2176, 2012.

S. L. Vehrencamp, J. Yantachka, M. L. Hall, and S. R. de Kort. Trill performance components vary with age, season, and motivation in the banded wren. *Behavioral Ecology and Sociobiology*, 67(3):409–419, 2013.

W. C. Warren, D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Kunstner, S. Searle, S. White, and others. The genome of a songbird. *Nature*, 464(7289):757–762, 2010.

T. Xu, W. Wang, and W. Dai. Sparse coding with adaptive dictionary learning for underdetermined blind speech separation. *Speech Communication*, 55: 432–450, 2013 .